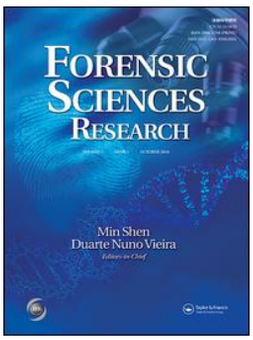
Quantitative evaluation of the results of digital forensic investigations: a review of progress

Journal Item



Forensic Sciences Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tfsr20

Quantitative evaluation of the results of digital forensic investigations: a review of progress

Richard E. Overill & Jan Collie

To cite this article: Richard E. Overill & Jan Collie (2021): Quantitative evaluation of the results of digital forensic investigations: a review of progress, Forensic Sciences Research, DOI: 10.1080/20961790.2020.1837429

To link to this article: https://doi.org/10.1080/20961790.2020.1837429

Published online: 08 Feb 2021.

FORENSIC SCIENCES RESEARCH
https://doi.org/10.1080/20961790.2020.1837429

REVIEW

OPEN ACCESS Check for updates

Quantitative evaluation of the results of digital forensic investigations: a review of progress

Richard E. Overill [a] and Jan Collie [b]

[a]Department of Informatics, King's College London, London, UK; [b]Department of Computing & Communication, The Open University, Milton Keynes, UK

ABSTRACT

Unlike conventional forensics, digital forensics does not at present generally quantify the results of its investigations. It is suggested that digital forensics should aim to catch-up with other forensic disciplines by using Bayesian and other numerical methodologies to quantify its investigations' results. Assessing the plausibility of alternative hypotheses (or propositions, or claims) which explain how recovered digital evidence came to exist on a device could assist both the prosecution and the defence sides in criminal proceedings: helping the prosecution to decide whether to proceed to trial and helping defence lawyers to advise a defendant how to plead. This paper reviews some numerical approaches to the goal of quantifying the relative weights of individual items of digital evidence and the plausibility of hypotheses based on that evidence. The potential advantages enabling the construction of cost-effective digital forensic triage schemas are also outlined.

KEY POINTS

- The absence of quantified results from digital forensic investigations, unlike those of conventional forensics, is highlighted.
- A number of approaches towards quantitative evaluation of the results of digital forensic investigations are reviewed.
- The significant potential benefits accruing from such approaches are discussed.

ARTICLE HISTORY
Received 17 April 2020
Accepted 11 October 2020

KEYWORDS
Forensic sciences; digital forensic metrics; measures of plausibility; Bayesian networks; probability theory; statistical theory; complexity theory; information theory

Introduction

One of the most striking differences between the results reported from conventional forensic investigations, involving the examination of physical, chemical and biological material traces on the one hand, and those from digital forensic investigations on the other, is the absence of any quantitative measures of confidence, plausibility or uncertainty associated with the latter results. To illustrate, the random match probability (RMP) of two matching DNA profiles not belonging to the same person (or to identical twins) is ca.10^{-8}, to within a factor of 10, depending on the number of loci in the profile and the size of population database [1]. Similarly, in forensic entomology examination of blow-fly larval instars can be used to determine the postmortem interval for a corpse, with a known range of uncertainty related to ambient temperature and humidity.

Such quantitative measures, generally derived the results of statistical analyses or laboratory experiments, are valuable since they enable both defence and prosecution sides to evaluate the strength (or weight) of an individual item of recovered evidence (e.g. a DNA profile) and, by extension, to estimate the strength (or plausibility) of a case built from many such evidential items.

This state of affairs is at least in part related to the relative maturity of conventional forensic science in comparison with digital forensics. We might tentatively trace the origin of systematic forensic science investigations to be ca.1900 with the publication of Edward Henry's fundamental work on fingerprints [2], followed in 1901 by the establishment of the Fingerprint Branch at New Scotland Yard. The subsequent enunciation by Edmond Locard of his well-known Exchange Principle that every contact leaves a trace [3] led to important conceptual and methodological advances in the science. Similarly, Cliff Stoll's tracking of the hacker Markus Hess [4] and Gene Spafford's decoding of the Robert Morris internet worm [5] could be taken as one measure of the beginning of systematic digital forensic investigations. It is immediately clear that conventional forensics has gained a head-start of around 90 years over digital forensics. However, given that digital forensic evidence is required to

CONTACT Richard E. Overill richard.overill@kcl.ac.uk

meet precisely the same admissibility criteria and tests of rigour in a court of law as conventional forensic evidence, it is apparent that there is a requirement to develop analogous quantifiable metrics for the findings of digital forensic investigations.

Fred Cohen, in particular, has made significant efforts to specify the rigorous scientific and engineering principles and practices upon which the requirements for such metrics should be based [6]. His work demonstrates that since individual binary bits have a physical instantiation it is possible to treat collections of them as a "bag of bits" using mathematical concepts from information physics, which can lead to quantitative findings. It is also worth noting that a recent Organization of Scientific Area Committees (OSAC) report [7] briefly considers the quantitative evaluation of investigative findings, and a Scientific Working Group on Digital Evidence (SWGDE) report [8] provides numerical error rates for some common digital forensic processes.

While there has been significant progress in specifying models and processes for the systematisation of digital forensic investigations, which are aimed at improving the consistency and reliability of the conclusions reached [9–15], it is important to emphasise that such developments, invaluable as they undoubtedly are within their own remit, do not attempt to directly address the issue of obtaining quantifiable findings from digital forensic investigations, analogous to those exemplified at the beginning of this section, which is the principal subject of this review article.

As a point of clarification, we should note here that the great majority of the work described in this review article refers to the quantification of hypotheses (propositions or claims) based on the digital evidence, rather than the quantification of the digital evidence itself; this is an important distinction. The only instances cited here involving the quantification of digital evidence itself occur in the examples where conditional probabilities (likelihoods) are assigned to the nodes of Bayesian networks based on surveys of experienced domain experts.

Proposition plausibility metrics

The formal relationship between plausibility and probability can be conveniently expressed as follows: probabilities signify the quantities that define a particular monotonic scale on which degrees of plausibility can conveniently be measured [16].

Bayesian methods, based on the conditional probability theorem of Revd. Thomas Bayes in his renowned posthumously published essay [17], have recently been cited as one approach to gaining quantitative traction in conveying degrees of (un)certainty in digital forensic results [18]. For a hypothesis (or proposition, or claim) H, with a single mutually exclusive and exhaustive alternative $\overline{H}$, and recovered evidence E, Bayes theorem can be conveniently expressed as:

$$\frac{\Pr(H|E)}{\Pr(\overline{H}|E)} = \frac{\Pr(H)}{\Pr(\overline{H})} \cdot \frac{\Pr(E|H)}{\Pr(E|\overline{H})}$$

where the left-hand side quotient represents the posterior odds ratio, and on the right-hand side the first quotient represents the prior odds ratio while the second quotient represents the likelihood ratio (LR). This simple expression can be generalised in a straightforward manner to situations involving multiple mutually exclusive and exhaustive alternative hypotheses.

Judea Pearl showed how networks can be defined and constructed which permit the propagation of probabilities from initial priors to final posteriors, based on the values of intervening conditional probabilities [19,20]. One of the first attempts to apply such quantitative methods to the analysis of an actual digital forensic investigation was made by Chow and co-workers using a Bayesian network model of an illicit peer-to-peer (BitTorrent) uploading case from Hong Kong, China [21]. The prior probabilities were taken to be strictly noninformative and the requisite conditional probabilities (likelihoods) were elicited from a survey of 31 domain experts. This model yielded a posterior probability of ca.92.5% in favour of the hypothesis that an illicit upload had indeed occurred given that all 18 anticipated items of digital evidence were recovered. Although a credible alternative hypothesis was not available for this case against which to compare the result, it corresponds to an LR of ca.12.3 in favour of the prosecution hypothesis. In subsequent studies, the sensitivity of this result to the absence of one (ca.0.08%), two (2.0%) or more items of digital evidence was found to be consistently small, while its sensitivity to uncertainties in the conditional probability (likelihood) values populating the nodes of the Bayesian network was also inconsiderable at ca.0.25% [22].

From 20 typical cases of internet auction fraud prosecuted in Hong Kong, China, Bayesian networks for both the prosecution and the defence cases were created and the LR of these two alternative explanations for the existence of the recovered digital evidence was computed to be 164 000 in favour of the prosecution hypothesis [23]. This finding may be interpreted as providing "very strong support" for the prosecution's hypothesis [24]. The conditional probabilities required for the Bayesian networks were sourced from a survey of the members of the Hong Kong Customs & Excise digital investigation team involved in the prosecutions. While LRs are generally regarded as the preferred

way to present forensic findings when at least two mutually exclusive and exhaustive hypotheses are available, it should nevertheless also be mentioned that there has been considerable debate regarding the possible (mis)interpretation of LRs potentially resulting in misleading conclusions being drawn [25].

A third example of Bayesian network analysis involves an actual case from Hong Kong, China of a leaked confidential (Yahoo!) email; the prior probabilities were taken to be strictly noninformative and the conditional probabilities (likelihoods) were elicited by questioning a domain expert. With every anticipated item of digital evidence successfully recovered the posterior probability in favour of the prosecution hypothesis was ca.97.2%. While a credible alternative hypothesis was not available for this case against which to compare the result, it corresponds to an LR of ca.34.7 in favour of the hypothesis; however, both single-parameter and multi-parameter sensitivity analyses resulted in minimal perturbations to that value [26,27].

While Bayesian networks deal mainly with conditional probabilities, these quantities can on some occasions be difficult to obtain in a reliable manner. In such situations it may instead be possible to apply conventional (frequentist) probability theory to the evaluation of the plausibility of a hypothesis put forward by either the prosecution or the defence side. For example, in cases where a seized computer is found to contain a relatively small number of illicit images or videos (e.g. of child pornography) amongst a much larger number of non-illicit material (e.g. of adult pornography) it is possible to use an Urn Model [28,29] and the Binomial Theorem to calculate the probability of the inadvertent download defence, under the assumption of random browsing activity. In two actual cases from Hong Kong, China the 95% confidence interval for the plausibility of this defence was shown to be ca.[0.03%, 2.54%] and ca.[0.00%, 4.35%], respectively [30].

In cases where very large quantities of illicit materials are recovered from a seized computer, the Trojan horse defence (THD) [31,32] is sometimes invoked to explain their presence. In such situations an analysis of the complexity of the processes involved in the alternative hypothetical explanations can be instructive. Operational complexity models count the number of operations (e.g. at byte-level) required to achieve the presence of the recovered materials by each of the two alternative mechanisms; the principle of least contingency, which asserts that (*ceteris paribus*) a simpler mechanism is more likely than a more intricate mechanism, then enables the odds ratio for the two alternatives to be computed as the inverse of the ratio of their complexities. In a particular scenario involving the deposition of a single 1MB image the odds against the THD were calculated to be just 2.979:1; these odds were lengthened to 197.9:1 if an up-to-date malware scanner was known to be operational at the material time [33]. A similar complexity based analysis has also been used to compute the odds against the THD for the five most commonly occurring cyber-crimes in Hong Kong, China [34], where ratios of between 100:1 and 1 000:1 were found; these odds ratios were subsequently furnished with worst case uncertainty bounds using standard error propagation theory to investigate the possibility of overlapping lower and upper bounds between the criminal and the THD hypotheses, respectively [35].

Knowing the relative plausibility of alternative hypotheses explaining the existence of the recovered digital evidence can be a valuable tool for aiding the investigating authority in deciding whether or not to refer a case to the prosecuting authority, and equally in assisting the prosecution authority in deciding whether or not to proceed to trial. Conversely, such quantitative information could also be made use of by the defence side in deciding how to plead: if the prosecution's case is overwhelmingly plausible then the defendant may be advised to plead guilty whereas if its plausibility is only marginal then a not guilty plea might be entered. In some jurisdictions it is also possible that an expert witness might be permitted to bring forward such data as part of their testimony at trial.

Probative value metrics

The probative value (or strength, or weight) of an individual item of digital evidence in the context of a particular criminal case reflects the degree to which the presence of that item of evidence, if recovered, contributes to the overall plausibility of the hypothesis concerning the processes that created all of the recovered digital evidence. Perhaps the simplest method to achieve this is to take the difference in the posterior probabilities of the Bayesian network for the hypothesis in the presence and in the absence of that item of evidence [36]. A second method is to generate the so-called Tornado diagram for the Bayesian network, which shows the ordered range of variation in posterior probability due to each evidential item with respect to all the remaining items [37]. Another, still more sophisticated, approach is to use information theory: the Kullback–Leibler divergence of the Shannon entropy gives the information gain as a measure of the difference between the probability distributions for the Bayesian network with (P) and without (Q) that item of evidence:

$$D_{KL}(P\|Q) = \sum_i P_i \log \frac{P_i}{Q_i}$$

These three approaches lead to somewhat different orderings for the evidential weights of the BitTorrent case mentioned above [38].

In digital forensic investigations it is commonly the case that the individual items of digital evidence may, at least to a first approximation, be considered conditionally independent of one another. Knowing the relative importance of each individual item of digital evidence in an investigation can then enable an efficient digital investigation scheme to be devised in which the most highly probative evidential items are searched for first, in order, relegating the items of lesser importance until later. If one or more of the anticipated items of high importance are not recovered the search may be de-prioritised or even abandoned; conversely, if all of the anticipated items of high importance are found then it may not be considered necessary to search for those items of lowest importance as the overall plausibility of the investigative hypothesis would not be sensibly improved by doing so.

In addition, such priority-driven investigation schemes can be termed cost-effective if they are extended to utilise economic criteria such as Return on Investment (RoI) or Cost–Benefit Ratio (CBR), through a knowledge of the resources (e.g. personnel, time, specialised equipment, etc.) required to recover and analyse each anticipated item of digital evidence [39,40].

Summary and conclusions

We have endeavoured to make the case that digital forensics should aim to catch-up with conventional forensics in providing quantified results from its investigations, despite the intrinsically sensitive and volatile nature of much digital evidence. A number of methodologies, such as Bayesian networks, complexity theory, probability theory and statistics, and information theory, by which this may be accomplished have been outlined and some typical results have been summarised. The uses to which such quantitative results could be put in both investigative and juridical contexts have also been briefly explored. In particular it should be emphasised that by bringing digital forensics into line with conventional forensic science, expert witness testimony and examiners' investigation reports can offer the courts a transparent rationale for the degree of certainty associated with their conclusions, rather than relying on previous relevant experience as the sole determinant of their expert opinion.

Ongoing and future lines of research in this area are likely to involve comparative studies of different measures of complexity [41] as applied to evaluating the plausibility of investigative hypotheses. Comparative studies of the divergence of various forms of entropy [42] applied to the quantification of evidential weight (or probative value) is also likely to be productive. Another potentially worthwhile avenue of research involves the pre-emptive analysis of the plausibility of novel, so-far-unused cybercriminal defences, for example the cosmic ray defence (CRD) where power law statistics guided by Moore's law enabled the prediction of a 512-fold increase in the incidence of CR-induced memory bit-flips since 1994, unless protected by sufficiently powerful error correcting codes [43].

Authors' contributions

Both authors, Richard E. Overill and Jan Collie, contributed to the final text of the review and approved it.

Compliance with ethical standards

The authors performed no studies involving human subjects or animals.

Disclosure statement

The authors disclose no conflicts of interest.

ORCID

Richard E. Overill http://orcid.org/0000-0002-5943-1812
Jan Collie https://orcid.org/0000-0002-3962-0173

References

[1] National Research Council. The evaluation of DNA forensic evidence. Washington, DC: NRC, National Academies Press; 1996. p. 34.
[2] Henry ER. The classification and use of fingerprints. London (UK): Government of India; 1898.
[3] Locard E. Manual of police techniques. 3rd ed, Pt. 1, Chapter III. Paris (France): Payot; 1939.
[4] Stoll C. Stalking the wily hacker. Commun ACM. 1988;31:484–497.
[5] Spafford E. The internet worm program: an analysis. SIGCOMM Comput Commun Rev. 1989;19:17–57.
[6] Cohen F. Digital forensic evidence investigation. 5th ed. Livermore (CA): ASP Press; 2013.
[7] Organization of Scientific Area Committees. A framework for harmonizing forensic science practices and digital/multimedia evidence, The Organization of Scientific Area Committees for Forensic Science, Technical Series 2. p. 7–8. 2019. Available from: https://www.nist.gov/sites/default/files/documents/2018/01/10/osac_ts_0002.pdf
[8] Scientific Working Group on Digital Evidence. Establishing confidence in digital and multimedia evidence forensic results by error mitigation analysis. Scientific Working Group on Digital Evidence, Appendix B; 2018. Available from: https://www.swgde.org/documents/Current.Documents/SWGDE
[9] Stephenson P. Structured investigation of digital incidents in complex computing environments. Inf Syst Secur. 2003;12:29–38.

[10] Stallard T, Levitt K. Automated analysis for digital forensic science: semantic integrity checking. Proceedings of the 19th Annual Computer Security Applications Conference (ACSAC'03); 2003 Dec 8–12; Washington, DC: IEEE Computer Society. p. 160–167.

[11] Gladyshev P, Patel A. Finite state machine approach to digital event reconstruction. Digital Invest. 2004;1:130–149.

[12] James J, Gladyshev P, Abdullah MT, et al. Analysis of evidence using formal event reconstruction. Proceedings of 1st International Conference on Digital Forensics and Cyber Crime (ICDF2C'09); 2009 Sep 30–Oct 2; Albany (NY). p. 85–98.

[13] Arasteh AR, Debbabi M, Sakha A. A formal approach for the forensic analysis of logs. Proceedings of the 5th conference on New Trends in Software Methodologies, Tools & Techniques (SoMeTT'06); 2006 Oct 25–27; Amsterdam (The Netherlands): IOS Press. p. 159–176.

[14] Arasteh AR, Debbabi M, Sakha A, et al. Analyzing multiple logs for forensic evidence. Digital Invest. 2007;4:82–91.

[15] Rekhis S, Boudriga N. Logic-based approach for digital forensic investigation in communication networks. Comput Secur. 2011;30: 376–396.

[16] Jaynes ET. Probability theory: the logic of science. Cambridge (UK): Cambridge University Press; 1995. Chapter 2.

[17] Bayes T. An essay towards solving a problem in the doctrine of chances. Phil Trans Roy Soc Lond. 1763;53:370–418.

[18] Casey E. Clearly conveying digital forensic results. Digital Invest. 2018;24:1–3.

[19] Pearl J. Reverend Bayes on inference engines: a distributed hierarchical approach. Proceedings of Natl Conf on A.I. (AAAI'82); 1982 Aug 18–20; p.133–136. Available from: https://aaai.org/Papers/AAAI/1982/AAAI82-032.pdf

[20] Pearl J. Probabilistic reasoning in intelligent systems. San Francisco (CA): Morgan Kaufmann; 1988.

[21] Kwan M, Chow, KP, Law, F, et al. Reasoning about evidence using Bayesian networks. In: Ray I and Shenoi S, editors. Advances in digital forensics IV. Boston (MA): Springer; 2008. Chapter 22; p. 275–289.

[22] Overill RE, Silomon JAM, Chow KP, et al. Sensitivity analysis of a Bayesian network for reasoning about digital forensic evidence, 4th International Workshop on Forensics for Future Generation Communication Environments. Proceedings of 3rd International Conference on Human-Centric Computing. Cebu (Philippines): IEEE Press; 2010 Aug 11–13; p. 228–232.

[23] Kwan YK, Overill R, Chow KP, et al. Evaluation of evidence in internet auction fraud investigations. In: Peterson G and Shenoi S, editors. Advances in digital forensics VI. Berlin (Germany): Springer; 2010. Chapter 7; p. 95–106.

[24] European Network of Forensic Science Institutes. ENFSI guideline for evaluative reporting in forensic science v3.0; 2015, p. 17. Available from: http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

[25] Morrison. Science & justice special virtual issue: measuring and reporting the precision of forensic likelihood ratios: introduction to the debate. Sci Justice. 2016;56:371–373.

[26] Kwan M, Overill R, Chow KP, et al. Sensitivity analysis of Bayesian networks used in forensic investigations. In: Peterson G and Shenoi S, editors. Advances in digital forensics VII. Berlin (Germany): Springer; 2011. p. 231–243.

[27] Overill RE, Zhang EP, Chow KP, et al. Multi-parameter sensitivity analysis of a Bayesian network from a digital forensic investigation. Proceedings of ADFSL Conference on Digital Forensics, Security and Law; 2012a May 30–31; Richmond (VA).

[28] George E. UK computer misuse act — the Trojan virus defence. Digital Invest. 2004;1:89–89.

[29] Bowles S, Hernandez-Castro J. The first 10 years of the Trojan Horse defence. Comput Fraud Secur Bull. 2015;2015:5–13.

[30] Overill RE, Chow KP. An approach to quantifying the plausibility of the inadvertent download defence. Forensic Sci Res. 2016;1:28–32.

[31] Bernoulli J. Ars conjectandi: Usum & applicationem praecedentis doctrinae in Civilibus. Moralibus & Oeconomicis [The art of conjecturing: use and application of the previous doctrine to civil. Moral and economic affairs]. Basel (Switzerland): Thurneysen Brothers; 1713. Chapter 4.

[32] Johnson NL, Kotz S. Urn models and their application: an approach to modern discrete probability theory. New York (NY): Wiley; 1977. Chapter 2.

[33] Overill RE, et al. Quantitative plausibility of the Trojan Horse Defence against possession of child pornography. Proceedings of 1st International Conference on Digital Forensics and Investigation (ICDFI 2012); 2012 Sep 21–23; Beijing (China). Available from: http://secmeeting.ihep.ac.cn/paper/Paper_Overill_Richard_ICDFI2012.pdf

[34] Overill RE, Silomon JAM. A complexity based forensic analysis of the Trojan Horse Defence. Proceedings of 4th International Workshop on Digital Forensics (WSDF 2011); 2011 Aug 22–26; Vienna (Austria). p. 764–768.

[35] Overill RE, Silomon JAM. Uncertainty bounds for digital forensic evidence and hypotheses. Proceedings of 5th International Workshop on Digital Forensics (WSDF 2012); 2012 Aug 20–24; Prague (Czech Republic). p. 590–595.

[36] Overill RE, Chow KP. Measuring evidential weight in digital forensic investigations: a role for Bayesian networks in digital forensic triage. In: Peterson G and Shenoi S, editors. Advances in digital forensics XIV. Cham (Switzerland): Springer; 2018. p. 3–10.

[37] Agena Ltd. AgenaRisk 7.0 user manual; 2016. p. 99–102. Available from: https://docplayer.net/57659618-Agenarisk-7-0-user-manual.html

[38] Schneps L, Overill R, Lagnado D, et al. Ranking the impact of different tests on a hypothesis in a Bayesian network. Entropy. 2018;20:856–870.

[39] Overill RE, Kwan YK, Chow KP, et al. A cost-effective digital forensics investigation model. In: Peterson G and Shenoi S, editors. Advances in digital forensics V. New York (NY): Springer; 2009. p. 193–202.

[40] Overill RE. Digital forensonomics — the economics of digital forensics. Proceedings of 2nd International Workshop on Cyberpatterns (Cyberpatterns 2013); 2013 Jul 8–9; Abingdon (UK).

[41] Lloyd S. Measures of complexity: a non-exhaustive list. IEEE Control Syst. 2001;21:7–8.
[42] Crupi V, Nelson JD, Meder B, et al. Generalized information theory meets human cognition: introducing a unified framework to model uncertainty and information search. Cogn Sci. 2018;42:1410–1456.
[43] Overill RE. Cosmic rays: a neglected potential threat to evidential integrity in digital forensic investigations? Proceedings of 15th International Conference on Availability, Reliability and Security (ARES 2020); 2020 Aug 25–28; Dublin (Ireland) (virtual event).